Using Inquiry-Based Methods to Teach Spatial Statistics: Decoupling Pattern From Process Using Bigfoot Sightings

ANDREW J. GREGORY[1,2] AND EMMA S. SPENCE[1]

[1]School of Earth Environment and Society, Bowling Green State University, Bowling Green, OH,USA
[2]Department of Biology, University of North Texas, Denton, TX, USA
Email: andrew.gregory@unt.edu

ABSTRACT Spatial statistics and experimental design are among the most important topics students in the environmental and ecological sciences learn and utilize throughout their careers. These topics are also among the most difficult for students to learn, often due to the use of contrived data sets that present simplified and unrealistic scenarios that fail to engage students in higher level thinking. One way to engage students in higher level thinking is to use an inquiry-based pedagogical framework. The use of inquiry as a pedagogical approach should be instinctive for most scientists, as it mimics how science is conducted, yet most instructors continue to use lecture-based, textbook-driven instructional formats. This type of approach is efficient in covering material, but it suffers in its ability to engage students or enhance learning. Using a Bigfoot data set in an inquiry-based framework, students in a cross-listed graduate/undergraduate statistics class learned ordinary least squares regression and geographically weighted regression techniques. These techniques are among the most frequently applied analyses in the natural sciences. The use of a Bigfoot data set engaged students' interest, rendering the prospect of learning regression topics as an emergent property of their interest and engagement. This approach also has an additional benefit in that students learned not only key statistical concepts but also learn how to self-diagnose deficiencies with their models as well as how to identify strategies to overcome these deficiencies. We hope that both instructors and students in graduate and undergraduate statistics or spatial modeling courses find this case study, and included data sets, a useful and interesting approach to teach and learn regression and spatial regression. **KEYWORDS** spatial statistics, regression, geographically weighted regression, Bigfoot

KEY MESSAGE

Readers of this case study will learn about experimental design; how to download data independently and then utilize and interrupt two of the important statistical techniques frequently used in the life sciences—ordinary least squares (OLS) regression and geographically weighted regression (GWR)—to draw conclusions about spatial patters and factors that are correlated to the "occurrence" of Bigfoot in various Ohio counties. We also cover some basics associated with spatial data analysis and the use of geographic information systems (GIS) software for data analysis. We necessarily assume that users of this case study already have a basic understanding of statistical data analysis and have some experience working with spatial data/GIS. For those lacking in these skills, or those in need of a refresher, there are a number of excellent resources available. For a review of general statistical concepts, we recommend *PDQ Stats* [1], *Experimental Design and Data Analysis for Biologists* [2], *Design of Experiments* [3], or *Biological Statistics: The Bare Essentials* [4]. For those wanting a review or primer of GIS concepts, we recommend *Statistical Analysis of Geographic Information* [5], *GIS, Spatial Analysis and Modeling* [6], or the ESRI training website [7]. Any one of these resources will provide users with an adequate refresher and/or review to make the concepts covered in this case study both easily accessible and useful. We have intentionally chosen general and broadly applicable sources to be referenced multiple times in order to reduce the necessity for readers to consult outside resources so as to make this case study applicable.

Case Studies in the Environment, 2020, pps. 1–10. electronic ISSN 2473-9510. © 2020 by the Regents of the University of California. All rights reserved. Please direct all requests for permission to photocopy or reproduce article content through the University of California Press's Reprints and Permissions web page, www.ucpress.edu/journals.php?p=reprints. DOI: https://doi.org/10.1525/cse.2020.1223915.

Key substantive content: How to design an experiment, download relevant data, and perform OLS regression and GWR in ArcMap [8, 9] and R [10].

Key messages: Spatial regression can be used to help identify the spatial dependence of patterns and processes to address environmental problems.

INTRODUCTION

Scientific inquiry presupposes that events in the universe occur in regular patterns that can be observed, quantified and then used to make predictions that can give us insights into underlying mechanisms [11]. One of the essential attributes of a good researcher, therefore, is curiosity. Cultivating this curiosity and encouraging a habit of careful questioning are both necessary components of teaching the scientific process [11]. However, most traditional (i.e., lecture or lab based) teaching modalities stifle the innate curiosity of students by providing highly contrived scenarios with an a priori "correct" solution to the laboratory [12]. Previous research has shown that these traditional teaching modalities result in student learning becoming less intrinsic (motivated by internal curiosity) and more extrinsic (motivated by a desire for instructor approval or a "*good grade*") [13–15].

Inquiry-based learning approaches are widely recognized as a way to remedy the above situation by harnessing student interest and curiosity to drive learning [16]. In a general sense, inquiry-based learning approaches are a form of experiential learning, but not all hands-on active learning practices are inquiry learning [17]. Good inquiry lessons will generally do the following and/or have the following attributes in common: (A) create an authentic experience, (B) are student-driven, (C) do not let the instructor be the expert, (D) the instructor asks questions without providing answers and/or there are no "correct" answers, (E) use humor to maintain student interest, and (F) stress not only the "know how" but also the "know why" of a particular lesson [16].

In other words, inquiry-based instruction is more constructivist, student-driven, and there are no a priori correct results [17, 18]. The focus of inquiry-based learning is on creative problem-solving, learning, and applying a specific methodology [12]. Therefore, the role of the instructor in an inquiry-based classroom is different from the role of the instructor in a conventional classroom: Instead of simply providing direct instruction (i.e., lecturing or guiding students through a lab exercise) to students, instructors act as a conduit to help students generate their own content-related questions and guide their investigation [12, 18]. In short, inquiry is a more natural and scientific way to teach burgeoning scientists how to use the tools of the trade.

In the environmental and ecological sciences, patterns can be used to understand disturbance processes or to delineate underlying mechanisms that are driving environmental heterogeneity and impacts. A variety of statistical tools and concepts have been developed to identify, tease apart, and test these patterns. One such statistical test is regression. Regression is used to evaluate relationships between two or more feature attributes or variables.

Identifying and modeling relationships allows one to better understand what is happening in a geographic location, predict where something is likely to occur, and begin to examine correlative relationships between events to better understand why things occur where they do [19]. Two of the most commonly applied regression methods are OLS and GWR. For OLS and GWR analyses, you need data sets that have: (1) a large number of observations, (2) a response factor that is heterogeneously distributed, and (3) a set of covariate (i.e., explanatory variables) variables that are distributed across the full range of the response variable.

To teach and/or learn OLS and GWR using injury-based approaches, Bigfoot-sightings data can be used, as they meet all of the requirements of inquiry-based learning (see above: A–F) and the data requirements of OLS and GWR analysis (see above: 1–3). It should be noted that the students in the following case study had previously been introduced to model selection using information theoretic approaches with AICc, hot spot analysis, OLS regression, and basic use of Program R. In addition, a course prerequisite was to have had previously taken an introductory course in GIS. Basic descriptions of these approaches are described in this case study; however, this case lesson is designed for implantation in an upper level undergraduate or graduate statistics or spatial modeling course, and it is therefore recommended that this lesson only be adapted for use in a classroom where the instructor is familiar and comfortable with regression analysis and an information theoretic approach to model selection. We provide a brief coverage of AIC/AIC_C model selection in a section below, but for those who might need a more thorough review of these concepts, see Norman and Streiner [4] for a review on regression

and Burnham and Anderson [20] for a review of information theory and model-selection procedures.

CASE EXAMINATION

Cryptozoology is a fringe discipline of biological inquiry where primarily hobbyists search for and study animals whose existence or survival is disputed or unsubstantiated. Bigfoot is among America's most legendary and charismatic cryptozoological species and, when studied, has always inspired more questions than answers [21]. The use of the name "Bigfoot" can be traced back to Northern California in 1958, where a construction worker, Jerry Crew, brought a plaster cast of a huge animal-like footprint, which he had found in the muck of Bluff Creek Valley, into a local newspaper office [22]. Nine years later, grainy video footage of what was claimed to be Bigfoot was shot by Roger Patterson and Robert "Bob" Gimlin in Bluff Creek [22]. Fast forward 60 years and Bigfoot is still being "spotted" across the United States. Such sightings have been documented by the Bigfoot Field Researcher Organization (BFRO) and cataloged in their online database [23].

OLS and GWR analyses can be applied to the BFRO and U.S. Census data sets to understand spatial patterns of where Bigfoot has been spotted and model the influence of covariates and space. By learning how to apply OLS and GWR techniques to understand drivers of Bigfoot sightings and spatial covariates, students can learn how to apply these tools to other data sets to ask similar questions.

OLS regression is among the best known and commonly applied of all regression techniques and is also the proper starting point for all spatial regression analyses. OLS provides a global model of the variable or process you are trying to understand or predict by creating a single regression equation to represent that process [4, 19]. However, because many environmental problems vary across the landscape (i.e., location and distance are a pivotal factor), it is often the case that OLS approaches fail to accurately model the data. OLS can, therefore, yield unreasonable or spurious results [8]. Another form of regression can be used to account for spatial variation: GWR. GWR is a statistical technique designed to deal with spatial heterogeneity in response factors across the landscape [8]. Instead of assuming that a single model can be fitted to the entire study region, it looks for geographical differences among regions. GWR works by moving a search window (called a kernel) from one point in a data set to the next, working through them sequentially. As the search window rests on a sample point, all other points encompassed by the kernel are identified and spatial autocorrelation is tested among all points within the kernel [8]. Spatial autocorrelation is a situation where you have high values of the response variable clustered near other high values and low values clustered near other low values.

There are two options for how the area encompassed by the kernel is identified: adaptive and fixed. In fixed kernel sampling, the number of points in the kernel will remain the same, but the spatial area covered may vary; in adaptive kernel sampling, the number of points and spatial area will vary, such that only kernels with a significant spatial autocorrelation will be retained. The choice of kernel estimation is an important consideration for GWR, but as a rule of thumb choose adaptive kernel sampling [8]. For more information on GWR and kernel estimation, see supplemental documents.

DATA SETS

Data sets have been provided in the supplemental data; however, we encourage readers, students, and instructors to augment the provided data sets with their own data to test variables of intrinsic interest to yourself or your students. Although we strongly encourage the use of data not presented in this case study, it should be noted that the use of new data will result in the code needing to be adapted for use with your data sets.

Response Variable (Bigfoot Sightings)

The BFRO's [23] online database contains a comprehensive county-specific list of Bigfoot sightings for the continuous United States and province-specific data for Canada. For the case study described here, we use BFRO data from Ohio.

Explanatory Variables

Since the Bigfoot data set is based on county, all explanatory variables must also be aggregated to the county level. In an inquiry framework, data should also be of interest to the individual conducting the study, for example, previous students have used number of black bear sightings per county, number of hunters per county, number of marijuana or methamphetamine arrests per county, or any other county-wide data that are available freely. For this case study, we use the U.S. Census Bureau data [24]. Included is a suitable data set of 47 explanatory variables

co-distributed with the Bigfoot data set (See supplemental materials for example datasets, "bfdata.csv"). For the purpose of the case study presented here, we chose to report variable names in our models and results as they are recorded in the U.S. Census Bureau database. Therefore, any racial or socioeconomic stereotypes implied by the models are anachronisms of the Census Bureau naming conventions.

ANALYSIS. To prepare students for work with GWR, it is necessary to give a brief primer of OLS regression, model selection (e.g., AIC), and hot spot analysis for estimating local spatial autocorrelation among covariates. Below, we provide a brief overview of each of these techniques prior to going over the details of our case study. Included with this section is necessary R code for duplicating the analyses presented in this case study.

Prior to starting your analysis, you will need to identify a region of interest (ROI) for your study. The ROI defines the spatial extent for your study and is an important attribute of the analysis as it directly impacts the validity of your results [25]. In general, the ROI should be defined by the spatial extent of your data sets, both the response factors and predictor covariates. In terms of this case study, the ROI was defined by the Bigfoot sightings and U.S. Census Bureau data collected by county for the state of Ohio. Extrapolations of model predictions beyond the ROI should be avoided if possible, as the validity of the results becomes more dubious the further outside the ROI you extrapolate [25].

OLS

The goal of linear regression is to model the relationship between a response variable and one or more explanatory variables (covariates). OLS regression is the most commonly applied linear regression technique [4]. The method uses the equation of a line to solve the data dependencies question among predictors and responses:

$$\hat{y}_{ij} = \beta_0 + \beta_1 X_{ij1} \ldots + \beta_n X_{ijn} + \hat{e}_{ij}$$

Where $\hat{y}_{ij}$ is the estimated value of the response variable for each observation i–j, β_0 is the intercept, $\beta_1 X_1 \ldots \beta_n X_n$ are predictor variables 1 through n multiplied by the original covariate variable for each observation i–j, and $\hat{e}_{ij}$ is the unexplained variance, calculated as the residual sum squares of the difference between the predicted value of the response variable ($\hat{y}_{ij}$) for each observation and the observed value of the response variable for each observation i–j. We call this term simply the residual [4, 19]. OLS is a computational approach that solves the linear model in such a way as to minimize the value of $\hat{e}_{ij}$, in essence to minimize the amount of unexplained variance [19, 26].

In the supplemental materials for this case study, we provide an annotated Rscript ("BF_OLS_CODE.R") that will allow the reader to run the OLS regression on the included data file ("bfdata.csv") necessary to obtain the results presented in this case study. The annotation in the Rscript describes what each line of code in the script does. Below is an example of the result of an OLS regression that models the effects of the number of male citizens per county ("MALES") and the number of ethnically white citizens in the county ("WHITE") on the number of Bigfoot sightings per county in 2015 ('X2015_BF'):

```
Call:
lm(formula = X2015_BF ~ MALES + WHITE, data = bfdata)

Residuals:
Min      1Q  Median      3Q     Max
-4.8073 -2.2592 -0.8597  1.9586 14.8301

Coefficients:
             Estimate   Std. Error  t value Pr(>|t|)
(Intercept)  1.630e+00  5.165e-01   3.156 0.002210 **
MALES       -1.584e-04  4.566e-05  -3.468 0.000825 ***
WHITE        9.885e-05  2.803e-05   3.527 0.000680 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.223 on 85 degrees of freedom
(1 observation deleted due to missingness)
Multiple R-squared:  0.1315,   Adjusted R-squared:  0.1111
F-statistic: 6.435 on 2 and 85 DF, p-value: 0.002499
```

In the above output from R, the p value of 0.002499 is the significance test that indicates that "MALES" and "WHITE" are significant predictors of "X2015_BF"; in other words, the number of White males in a county can accurately predict the number of Bigfoot sightings per county. The multiple R^2 value of 0.1315 suggests that this model can explain about 13.15% of the variance in Bigfoot sightings. The estimate is the β values for the intercept and each covariate ("MALES" and "WHITE"), and the F-statistic is another metric of overall model fit based on the F-static common to an analysis of variance. We refer the reader to Package glm2 user's guide [26] and Norman and Steiner [4] for further reference.

HOT SPOT ANALYSIS

A hot spot analysis first identifies regions of high spatial autocorrelation and then tests whether the regions are of high × high (hot spot) or low × low (cold spot) spatial

autocorrelation [9]. The result of the analysis is a map of where higher (hot spot) or lower (cold spot) than average measured values of a parameter are clustered together on the landscape. Hot and cold spots can be used to identify parameters (either response variables or covariates) that are distributed in a heterogenic pattern. If either hot spots or cold spots are detected for a particular parameter, then that suggests that a spatially varying regression model (i.e., GWR) is likely an appropriate statistical technique to apply to the data. If no spatial heterogeneity is detected (no hot spots or cold spots), then a simple OLS linear regression model is likely sufficient.

GWR

The key difference between GWR and OLS is that OLS fits a global regression analysis independent of local spatial autocorrelation, while the goal of GWR is to identify the spatial extent of spatial autocorrelation and fit localized regression functions across observations that are spatially correlated and then identify which variables are driving spatial heterogeneity and pattern [8].

The GWR form of this analysis modifies the generalized form of the of regression model by adding a spatial weights component to the model "M," which is essentially the strength of the spatial autocorrelation of each parameter "X" in the original model. This necessitates creating multiple new terms to the model to fit each potential interaction:

$$\hat{y} = \beta_0 + [\beta_1 * X_1] + [\beta_1 * X_1 * M_1] + [\beta_2 * X_2]$$

$$+[\beta_2 * X_2 * M_2] \ldots [\beta n * X_n] + [\beta n * X_n * M_n] + \hat{e}_{ij}$$

There are numerous examples where the global model of predictors of a response factor may give misleading results. In the included "Intro to GWR.pptx," we give the example of an OLS regression model used to predict drivers of housing prices in and near New York City in the United States (See Supplemental Material). For this example, we chose two houses, one in Queens and one in Cornwall, New York, which were listed for sale on Zillow.com in 2013. These two homes are located ~50 linear miles or 65.4 highway miles apart from each other on the landscape. To construct a linear regression, the list price ("COST") was input as the response variable. Predictor variables were the selected: lot size ("ACRES"), number of bedrooms ("BED"), number of bathrooms ("BATH"), the year the house was built ("YEAR"), and the square footage of the house ("SIZE"). For the overall model of: ("COST"~ "ACRES"+ "BED" + "BATH" + "YEAR" + "SIZE"). The result of this model was a significant regression that yielded a ridiculous result such that being homeless would cost around $10M, but someone would actually pay you to live in a 12 bedroom, 8 bathroom mansion. This example from the supplementary material highlights the key problem with OLS global regression approaches, namely that they sometimes yield very unreasonable results. This is particularly true if there is latent spatial dependency in the response variables (i.e., there is high spatial autocorrelation among responses [8]).

In the case of the homes for sale in New York, it is obvious why this is likely to occur; homes in New York City are much more expensive than homes in the New York countryside. However, in other instances, the drivers of spatial variability in response factors may be less obvious. GWR can be implemented from a stand-alone graphical user interface, GWR4 [27]; as a Package in Program R, glm2 [10, 26]; or by using the GWR tool in ArcMap [28].

MODEL SELECTION USING AKAIKE'S INFORMATION CRITERION (AIC/AIC$_C$)

The goal of model selection is to identify a model or suite of potential good models that describe the system being modeled in as few parameters as possible while still maintaining model predictive power (i.e., parsimony [20]). Information theoretic approaches to model selection have become mainstays in environmental and ecological sciences because they balance the explanatory power of a model with the simplicity of the model structure [20]. AIC is the Akaike information criterion, which is Hirotugu Akaike's method of solving the negative log likelihood function of the probability that the model is an adequate explanation of the observed phenomena (i.e., it's a linear estimate of how much information is lost using the fitted model as the explanation of the phenomena being studied [27]). AIC can also be corrected for small sample size through the use of AIC$_C$. In general, low AIC values indicate a better fit for the data and stronger evidence for the preference of one model over a model with a larger AIC value [20]. There are no fixed rules for how much lower an AIC value needs to be for the model to be chosen as a more parsimonious explanation, but one general rule of thumb with the use of GWR is that if the

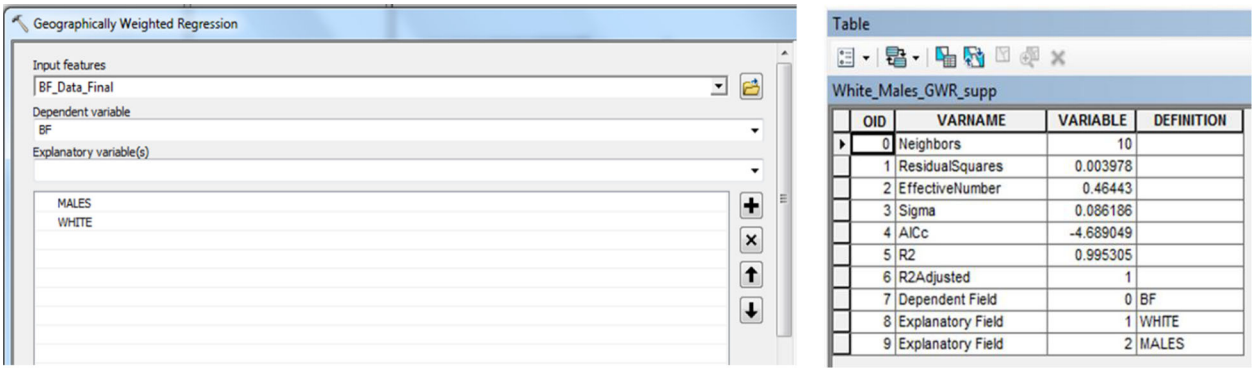

FIGURE 1. Example GWR and GWR output from a geographically weighted regression analysis implemented in ArcMap to model the ability of "WHITE" and "Males" to predict Big Foot sightings in Ohio. Note the R^2 value of 0.995 and the AIC_C value of −4.67.

GWR model represents a reduction in AIC of ≤ 4 (compared to other models, such as the same variables in an OLS model), then that generally suggests that the inclusion of a spatially heterogenic factor is a significant improvement of model performance, and so the GWR results should be used [8].

CASE STUDY USING BFRO AND CENSUS BUREAU DATA

Below you will find one student's results and how they used OLS and GWR techniques to model spatially explicit drivers of Bigfoot sightings on a county-by-county basis in Ohio. The students in this class were attending Bowling Green State University, located in Ohio, making it an ideal ROI for the class. Incidentally, at the time of this study, Ohio had the third highest number of Bigfoot sightings in the United States.

Drivers of Bigfoot sighting in Ohio

DATA. For the response variable of the regression, the BFRO's sightings for the state of Ohio (257 sightings in 88 Ohio counties) were used [23]. For the suite of explanatory variables, the 2010 U.S. Census Data were used to create a set of 47 potential explanatory variables [25].

METHODS. Data were first loaded into ArcMap and explored so that general trends could be observed ([28], figure 1). Prior to spatial data analysis with regression, an exploratory hot spot analysis in ArcMap was performed to explore possible spatial heterogeneity in the distribution of the response variable ([28], figure 2). If no spatial heterogeneity in the response variable exists, there is little reason to apply a GWR to the data set. The hot spot analysis showed that there was a cold spot in Bigfoot sighting in northwest Ohio and a hot spot in northeast Ohio ([28] figure 3).

After the dependent variable data were explored, the explanatory data set was accessed, and the dimensionality was reduced using a correlation matrix by running a Pearson's Product Moment Correlation analysis. All potential predictor variables correlated at $r \geq 0.25$ were removed from further consideration as potential explanatory covariates. This is a field standard, as closely correlated variables do not offer further explanatory power.

When eliminating variables, you have a choice as to which correlated variables to retain and which to remove. There are generally three criteria to be used when determining which variables to retain. First, you should retain all variables of primary interest to your study. For example, if you are primarily interested in how agricultural production influences Bigfoot sightings, you should retain all, or as many, variables related to agricultural production as possible. Second, you should use your understanding of system dynamics. If a particular variable is of primary importance to system function and another is superfluous, retain the key variable and eliminate the superfluous one(s). Third, data mine by conducting a series of univariate regression analyses and retain the variable that is most strongly correlated with the response factor (highest R^2 value). It should be noted that there is some debate among statisticians as to the validity of this last variable retention criteria. Some statisticians will argue that

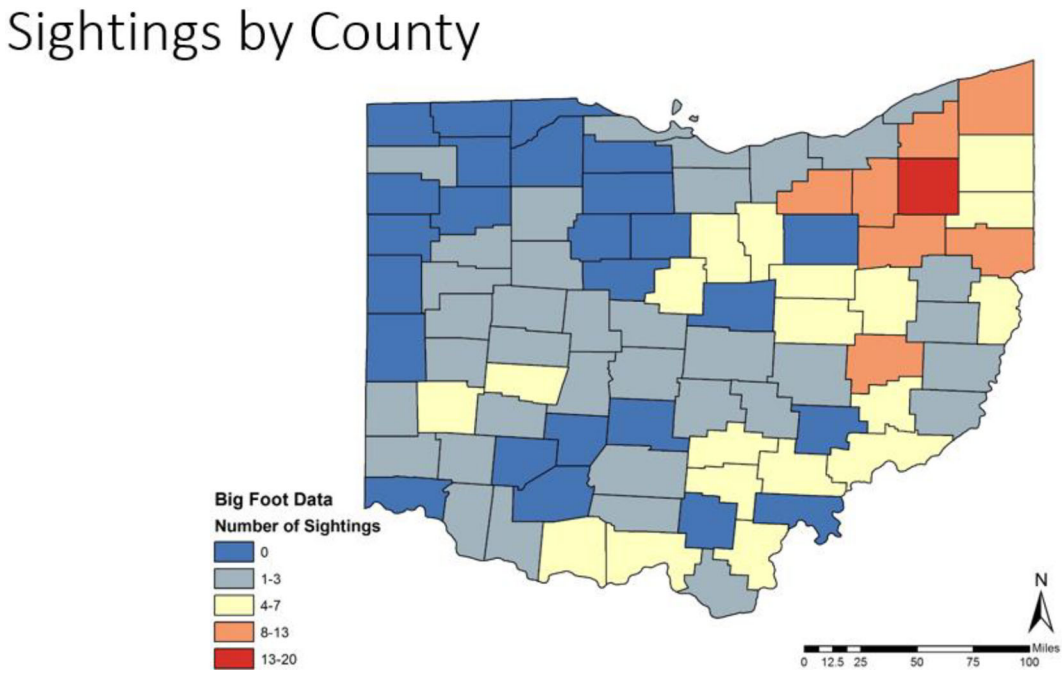

Figure 2. BFRO sighting data for Ohio counties based on frequency. Initial observations showed that there was a cluster of counties with a high number of sightings in the northeast. To see whether this is a significant trend, a hot spot analysis was conducted.

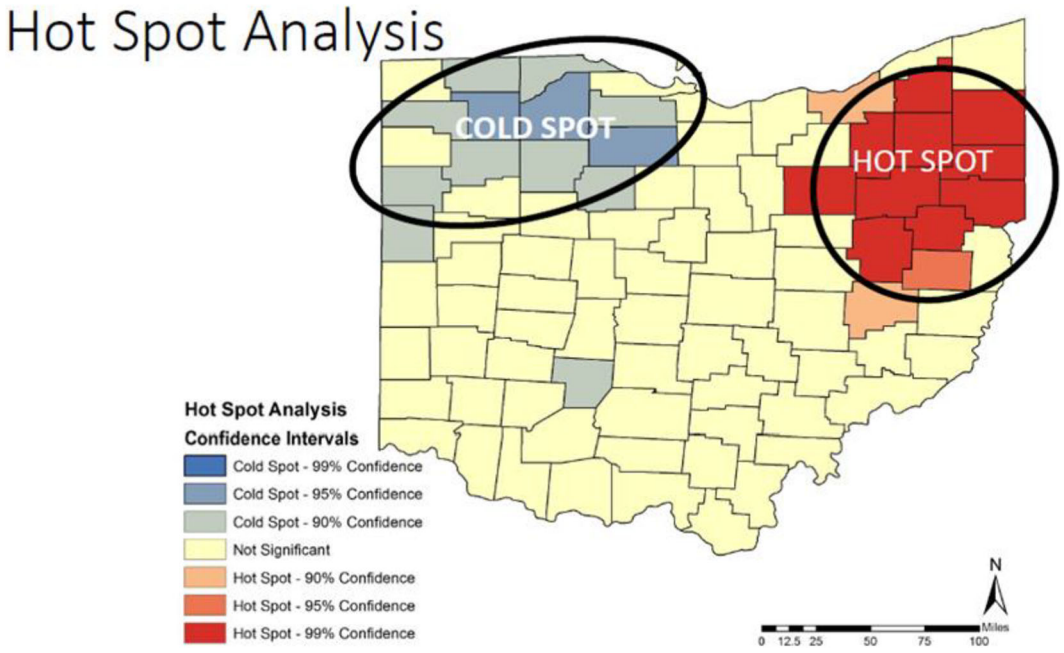

Figure 3. Hot spot analysis for Big Foot sighting by county in Ohio. Red and blue counties indicate whether the observed spatial clustering of high or low values is more pronounced than one would expect in a random distribution of those same values. Yellow counties are not significantly clustered.

such data mining is strictly forbidden and should never be done [2], and others will argue that data mining of this type is a best practice in statistical model development [1, 4, 20]. The authors of this article and case study fall into the latter camp, but readers should be aware of this ongoing statistical debate.

RStudio and ArcMap were used to run OLS regression, spatial regression, and GWR [10, 27]. An arbitrary total of 50 different models were tested with the number of input explanatory variables ranging from 1 to 15. An information theoretic approach to model selection based on AIC_C difference minimization was used to determine the most parsimonious model. Model selection via the AIC_C is able to determine whether explicitly modeling spatial heterogenic Bigfoot responses to the covariates increase model fit/performance or whether the perceived model improvement was due to uncontrolled variance inflation associated with the increased number of parameters needed to fit the spatial weights matrix to the data set [8].

Following GWR analysis, a spatial autocorrelation analysis was conducted using Moran's *I* in ArcMap on the GWR residuals to test for any remaining unexplained spatial variance. This serves as a post hoc tests of GWR model adequacy.

RESULTS

The most parsimonious model included the explanatory variables of (1) number of people that identified as Caucasian ("WHITE"), (2) number of mobile homes ("MOBILE"), and (3) number of individuals living in the county that are between 18 and 29 years old ("AGE_18_29"). Of the two regression techniques applied to the data sets, the GWR was consistently most parsimonious (table 1). Standardized residuals of the GWR were not spatially autocorrelated and were able to accurately predict Bigfoot sightings in 60 of the 88 Ohio counties ($R^2 > 0.6$), accounting for ~70% of the counties in Ohio. Other tested models (data not presented) included positive associations with number of males and a negative response to acres of cultivated agriculture per county.

These results also make sense, given the results of the spatial hot spot analysis. The cold spot found in northwest Ohio is found in a topologically homogenous region of Ohio with small isolated reserve areas surrounded by intense cultivated agriculture or embedded in urban centers such as Toledo. Conversely, the hot spot in northeastern Ohio is near Cuyahoga Valley National Park, which is characterized by more rugged terrain unsuitable for cultivation and a much higher density of wooded reserve areas sparsely populated with hunting cabins and vacation rental properties. Regardless, as no landscape attribute data were in the most parsimonious model, our data suggest that overall Bigfoot sightings are more likely observer driven than landscape or habitat driven.

TABLE 1. Results of the Three Different Types of Regressions Performed on the Most Parsimonious Linear Regression Model.

	Attribute	AICc
Linear regression	Non-Spatial	209.1
Spatial regression	Systematic spatial patters	464.36
GWR	Accesses localized patterns	−333.69

CONCLUSION

Although Bigfoot is America's most legendary creature and one that has always inspired more questions than answers [29], using OLS and GWR, we are able to model some potential explanatory variables that correlate to the areas that Bigfoot has been sighted while learning regression and model selection techniques. Just like many ecological phenomena, a mythical creature such as Bigfoot is not directly observable, so we have to use indirect evidence and models to support our conclusions. The BFRO online database can be used to create an inquiry-based lesson. Through this case study, the following learning objectives should have been met:

1. Students will be able to ask scientifically reasonable and tractable questions associated with the supplied data sets and will be able to articulate and justify their questions as well as the validity of their questions.
2. Students will be able to correctly apply OLS and GWR regression techniques to data and interpret the results both in a quantitative and qualitative manner.
3. Students will be able to identify limitations to the data sets they have and find creative solutions to solve analytical or design problems associated with the application of OLS and GWR to answer their chosen questions.
4. Students will be able to evaluate the suitability of an approach and validate their results.
5. Students will be able to justify the appropriateness of their analysis and experimental design choice and articulate their methods and results in a scientific manner.

CASE STUDY QUESTIONS

1. What are the uses of OLS and GWR regression approaches in environmental sciences?
   a. Come up with specific examples
2. What does the R^2 value tell us about model fit?
   a. What was the difference in R^2 between the model "BF ~ MALES + WHITE" and "BF ~ MALES + MOBILE + AGE_18_29"?
3. How can you model the influence of spatially heterogenic processes on a response factor?

AUTHOR CONTRIBUTIONS

All persons who met the authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript.

ACKNOWLEDGMENTS

The authors thank the students of the Quantitative Methods and Landscape Ecology classes for their hard work on the "Bigfoot Assignment." The authors also thank the Fisheries Blog at www.FisheriesBlog.net for publishing a blog about this lesson and providing the inspiration to write this article and share these data sets. Similarly, the authors thank the 300+ colleagues who, after reading the blog post, have contacted us for access to these data sets and further description of our pedagogical approach, thereby providing us the motivation to complete work on this article. We also thank the anonymous reviewers whose comments and critiques greatly improved our presentation of this article. Finally, we thank our respective institutions for allowing us the time and flight-of-fancy to pursue academic discourse on Bigfoot.

COMPETING INTERESTS

The authors have declared that no competing interests exist.

SUPPLEMENTAARY MATERIAL

All data sets used to create the included Bigfoot case study have been included in the supplemental materials, as well as teaching power point presentations to assist in the introduction of the lesson.

REFERENCES

1. Norman, GR, Streiner, DL. PDQ: Pretty Darned Quick Statistics. 3rd ed. London, UK: BC Dekker Inc; 2003.
2. Quinn GP, Keough MJ. Experimental Design and Data Analysis for Biologists. Cambridge, UK: Cambridge University Press; 2002.
3. Kuehl RO. Design of Experiments: Statistical Principles of Research Design and Analysis. 2nd ed. Albany, NY: Duxbury Press; 2000.
4. Norman GR, Streiner DL. Biostatistics: The Bare Essentials. 4th ed. Shelton, CN: Peoples Medical Publishing House; 2014.
5. Wong DWS, Lee J. Statistical Analysis of Geographic Information with ArcView GIS and ArcGIS. Hoboken, NJ: John Wiley & Sons; 2005.
6. Maguire DJ, Batty M, Goodchild MF. GIS, Spatial Analysis, and Modeling. Redlands, CA: ESRI Press; 2005.
7. ESRI. Training [Internet]. Redlands, CA: ESRI Press; 2017. Available: www.esri.com/training. Accessed 14 May 2017.
8. Fotheringham A, Brunsdon C, Charlton M. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. London, UK: John Wiley & Sons; 2002.
9. Mitchel A. The ESRI Guide to GIS Analysis Volume 1: Geographic Patterns and Relationships. Redlands, CA: ESRI Press; 2005.
10. R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing [Internet]. Vienna, Austria; 2018. Available: http://www.R-project.org. Accessed 14 May 2018.
11. American Association for the Advancement of Science [AAAS]. Science for all Americans [Internet]. Washington, DC: AAAS; 2018. Available: http://www.project2061.org/publications/sfaa/. Accessed 14 May 2018.
12. Crawford BA. Learning to teach science as inquiry in the rough and tumble of practice. J Res Sci Teach. 2007;44: 613–642.
13. Pugliese AC. The meaning of inquiry, discovery, and investigative approaches to science teaching. Sci Teacher. 1973; 39: 24–27.
14. Bianchini JA, Colburn A. Teaching the nature of science through inquiry to prospective elementary teachers: a tale of two researchers. J Res Sci Teach. 2000;37: 177–209.
15. Colburn, A. An inquiry primer. Sci Scope. 2000;23: 42–44.
16. Littleton K, Scanlon E, Sharples M. Orchestrating Inquiry Learning. New York, NY: Routledge Publishing; 2012.
17. White B. Enabling students to construct theories of collaborative inquiry and reflective learning: computer support for metacognitive development. Int J Artif Intell Educ. 1999;10: 151–182.
18. Lambert J. Using model-centered instruction to introduce GIS in teacher preparation programs. J Geosci Educ. 2007; 55: 387–395.
19. Weisberg S. Applied Linear Regression. Hoboken, NJ: John Wiley & Sons; 2005.

20. Burnham, KP, Anderson DR. Model Selection and Multimodel Inference: A Practical Information Theoretic Approach. 2nd ed. New York, NY: Springer-Verlag; 2002.
21. A&E Television Networks, LLC. Bigfoot Captured [Internet]. New York, NY: A&E Television Networks, LLC; 2018. Available: https://www.history.com/specials/bigfoot-captured. Accessed 14 May 2018.
22. Coleman L, Clark J. Cryptozoology A to Z: The encyclopedia of Loch Monsters, Sasquatch, Chupacabras, and Other Authentic Mysteries of Nature. New York, NY: Simon and Schuster Publishers;1999.
23. The Bigfood Field Researchers Organization. Main Page. BFRO; 2017. Available: www.bfro.net. Accessed 14 May 2018.
24. United States Census Bureau. American Community Survey. United States Census Bureau; 2018. Available: https://www.census.gov/acs/www/data/data-tables-and-tools/data-profiles/2015/. Accessed 14 May 2018.
25. Randin C, Dirnbrock T, Dullinger S, et al. Are species distribution models transferable in space? J Biogeogr. 2006;33(10):1689–1703.
26. Marschner IC. glm2: fitting generalized linear models with convergence problems. R J. 2011;3(2): 12–15.
27. National Centre for Geocomputation. Geographically Weighted Modeling. 2018. Available: http://gwr.maynoothuniversity.ie/gwr4-software/. Accessed 14 May 2018.
28. ESRI. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute; 2014.
29. Foxtopus Inc. Wild Thing [Internet]. Foxtopus; 2020. Available: https://www.foxtopus.ink/wildthing. Accessed 22 September 2020.